Universitas Indonesia Library >> UI - Skripsi Membership

Implementasi dan analisis efisiensi beberapa algoritma kompresi teks terhadap aspek-aspek informasi tekstual

Rian Yusran, author

Deskripsi Lengkap: https://lib.ui.ac.id/detail?id=125097&lokasi=lokal

---

Abstrak

Pada tugas akhir ini akan dibahas tinjauan mengenai beberapa teknik kompresi teks dan pengaruhnya terhadap bermacam sumber informasi. Algoritma kompresi yang akan dibandingkan adalah algoritma fano, shannon, huffman, alphabetic, arithmetic, lempel-ziv dan LZW. Semua algoritma tersebut penulis implementasikan dengan menggunakan bahasa pemrograman java. Algoritma ini diterapkan pada bermacam sumber informasi, dan dibandingkan pengaruh masing-masing algoritma kompresi terhadap sumber informasi tersebut. Sumber informasi yang dimaksud dalam skripsi ini adalah penghasil berita-berita berbahasa Indonesia yang dikelompokkan kedalam aspek ekonomi, politik, hankam, hukum, iptek, senibudaya, olahraga, dan kesehatan. Kedelapan macam sumber informasi itu dianalisis dengan metode sampling, yaitu dengan mengambil sampel berupa beberapa file berita. File-file berita tersebut diambil dari berbagai koran dan majalah online dalam rentang waktu dari 22 juni sampai 25 juni 2004 dan dari 4 agustus sampai 6 agustus 2004. Masing-masing aspek berita informasi ini akan mempunyai karakteristik yang berbeda-beda, sehingga perngaruh kompresi terhadap aspek berita tersebut berbeda-beda juga. Ada beberapa hal yang akan dibandingkan antara lain: compression ratio, efisiensi, kekhasan masing-masing metode kompresi. Compression ratio menunjukkan perbandingan besar file aslinya dengan file hasil kompresinya. Efisiensi menunjukkan perbandingan antara entropi dan panjang rata-rata untuk satu karakter. Hasil analisis terhadap ke-delapan macam sumber informasi tersebut menunjukkan bahwa tingkah-laku masing-masing sumber informasi tidak jauh berbeda satu sama lainnya. Untuk semua sumber informasi tersebut terlihat bahwa metode kompresi dictionary (universal source encoding) lebih baik dalam hal pemampatan bila dibandingkan dengan metode symbolwise.